VCEMATHSMETHODS.COM

MATHS METHODS

PODCAST

Sampling & populations

- Sample proportions
- Sampling distribution - small populations
- Sampling distribution - large populations
- Sampling distribution - normal distribution approximation
- Mean & variance of a sample proportion
- Sampling distribution - comparing approximations
- Mean & variance of the sample proportion
- Confidence intervals
- Margin of error

Sample proportions

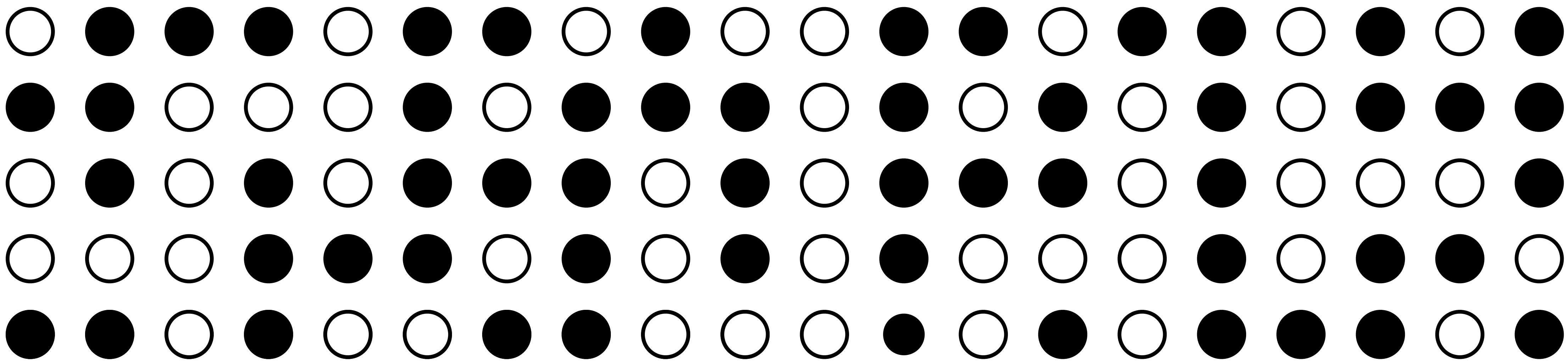
- A **sample** of size n is taken from a population.
- The number of positive outcomes in the sample is recorded to find the **sample proportion**.
- The **population proportion** can be estimated from the sample proportion.
- The sample proportions $\hat{p}$ are the values of the random variable $\hat{P}$.

$$p = \frac{\text{number of positive outcomes in population}}{\text{population size}}$$ (A population statistic.)

$$\hat{p} = \frac{\text{number of positive outcomes in sample}}{\text{sample size}}$$ (A sample statistic.)

$\hat{P}$ = The set of possible outcomes of $\hat{p}$.

Sample proportions



(Black is the positive outcome here)

$$p = \frac{\text{number of positive outcomes in population}}{\text{population size}} \qquad p = \frac{54}{100} = 0.54$$

$$\hat{p} = \frac{\text{number of positive outcomes in sample}}{\text{sample size}} \qquad \hat{p} = \frac{5}{10} = 0.5$$

$$\hat{P} = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$$

Sampling distribution - small populations

- If a population is small, then the probability of a selection changes depending on the previous selections. **(Conditional probability.)**
- For example, a group of 5 students is to be randomly selected from 12 boys and 10 girls.
- What is the **sampling distribution** for the proportion of boys selected?

$$Pr(X=x)=\frac{{}^{D}C_{x}\ {}^{N-D}C_{n-x}}{{}^{N}C_{n}}$$

$$Pr(X=2)=\frac{{}^{12}C_{2}\times{}^{10}C_{3}}{{}^{22}C_{5}}$$

$$Pr(X=2)=0.3008$$

$$Pr(\hat{P}=0.4)=0.3008$$

This is known as a **hypergeometric** distribution.

$Pr(\hat{P}=0)$	$=0.0096$
$Pr(\hat{P}=0.2)$	$=0.0957$
$Pr(\hat{P}=0.4)$	$=0.3008$
$Pr(\hat{P}=0.6)$	$=0.3759$
$Pr(\hat{P}=0.8)$	$=0.1880$
$Pr(\hat{P}=1)$	$=0.0300$

Sampling distribution - large populations

- If a population is sufficiently large, the probability of selection remains constant. **(Independent probability.)**
- For example, a group of 5 students is to be randomly selected from a large population at the school. (1000+ students, where 6/11 of the students are boys and 5/11 girls.)

$$Pr(X=x)={}^{n}C_{x}\times(p)^{x}\times(1-p)^{n-x}$$

$$Pr(X=2)={}^{5}C_{2}\times\left(\frac{6}{11}\right)^{2}\times\left(\frac{5}{11}\right)^{3}$$

$$Pr(X=2)=0.2794$$

$$Pr(\hat{P}=0.4)=0.2794$$

This is known as a **binomial** distribution.

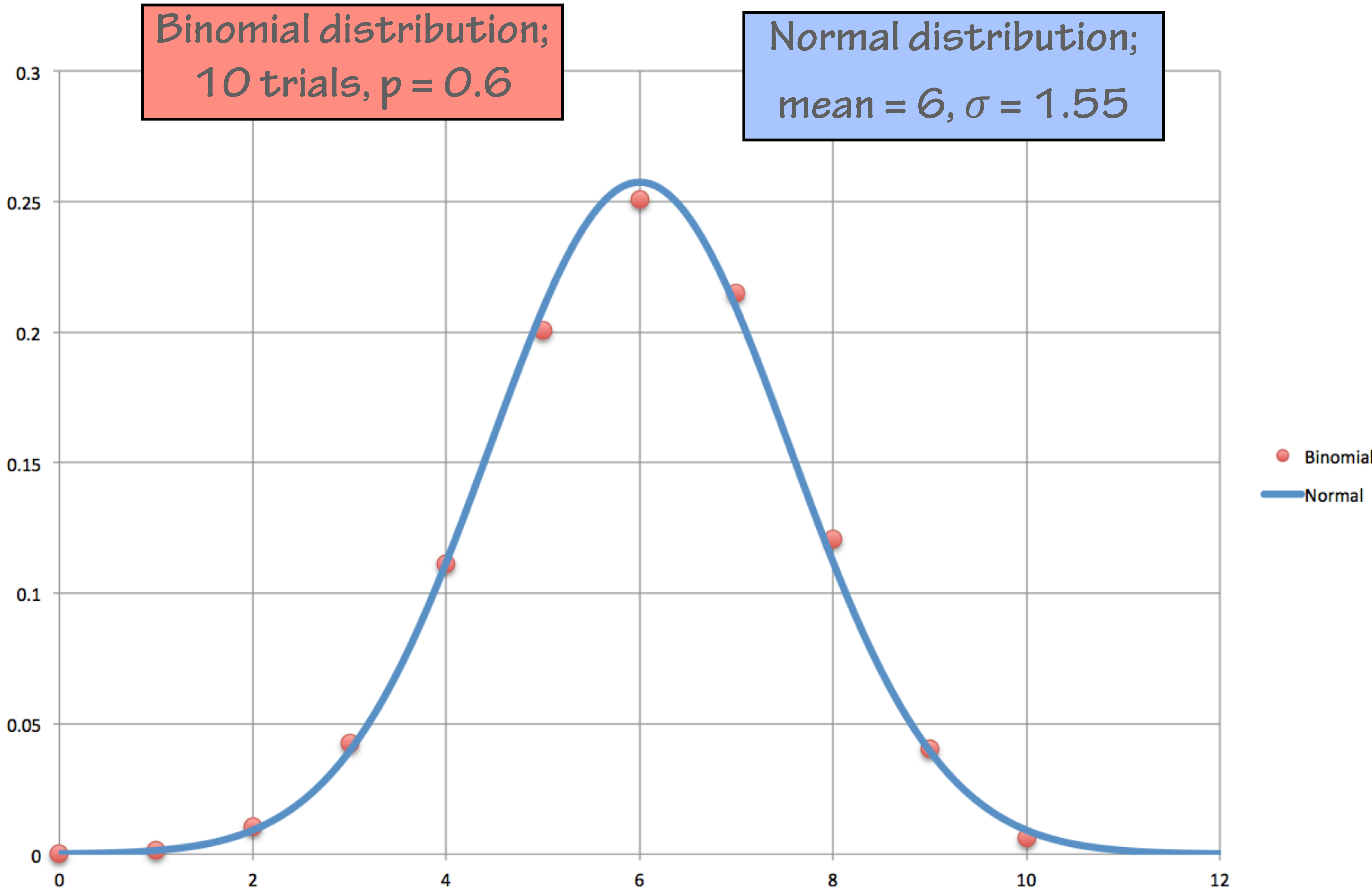
$Pr(\hat{P}=0)$	$=0.0194$
$Pr(\hat{P}=0.2)$	$=0.1164$
$Pr(\hat{P}=0.4)$	$=0.2794$
$Pr(\hat{P}=0.6)$	$=0.3353$
$Pr(\hat{P}=0.8)$	$=0.2012$
$Pr(\hat{P}=1)$	$=0.0483$

Sampling distribution - normal distribution approximation

- If a population is sufficiently large and the value of p is not too far from 0.5, the binomial distribution can be approximated by a normal distribution
- The binomial mean and standard deviation can be used with a normal distribution.
- For a binomial distribution:

$$\mu = E(X) = np \qquad sd = \sqrt{np(1-p)}$$

Sampling distribution - normal distribution approximation



Mean & variance of a sample proportion

- If a sample of n is taken from a population with a proportion p:

$E(X) = np$ (Binomial Mean)

$E(\hat{P}) = E\left(\frac{X}{n}\right)$

$E(\hat{P}) = p$

$Var(X) = np(1-p)$ (Binomial Variance)

$Var(\hat{P}) = Var\left(\frac{X}{n}\right)$

$Var(\hat{P}) = \frac{1}{n^2}Var(x)$

$Var(\hat{P}) = \frac{p(1-p)}{n}$

The expected value of the sample distribution is: $E(\hat{P}) = p$

The standard deviation of the sample distribution is: $sd(\hat{P}) = \sqrt{\frac{p(1-p)}{n}}$

Sampling distribution - comparing approximations

- 60% of people in a town are overweight. If a group of 100 people was to be randomly selected for a health survey, what is the probability that less than 55% of those surveyed are overweight?

- Binomial distribution: $\Pr(\hat{p} < 0.55)$

$$\Pr(0 < x < 54)$$

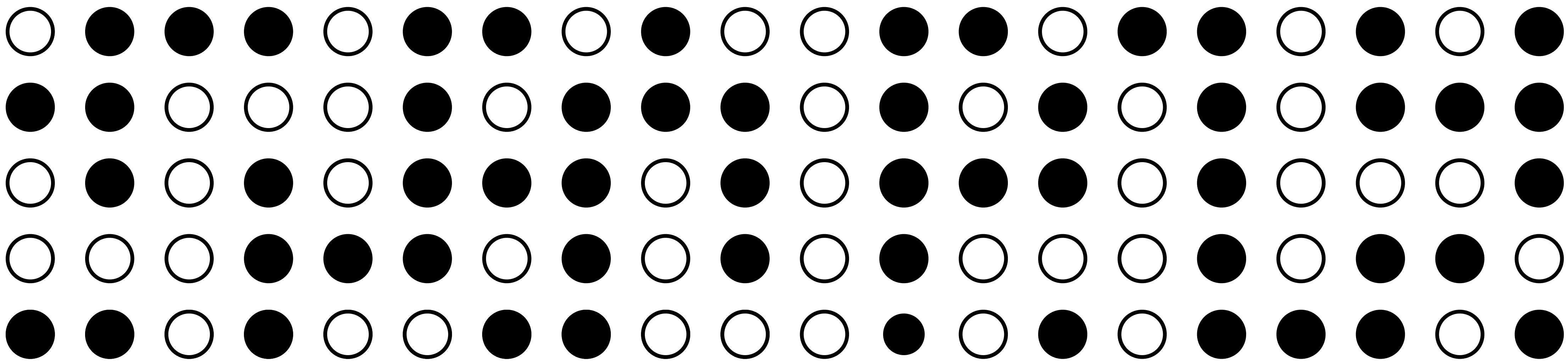
$$\text{binomCdf}(100, 0.6, 0, 54) = 0.1311$$

- Normal distribution: $\mu = E(\hat{p}) = 0.6$

$$\sigma = \sqrt{\frac{0.6 \times 0.4}{100}} = 0.0490$$

$$\text{normCdf}(-\infty, 0.55, 0.6, 0.0490) = 0.1537$$

Sample proportions



$$p = \frac{\text{number of positive outcomes in population}}{\text{population size}}$$

$$\hat{p} = \frac{54}{100} = 0.54$$

Mean & variance of the sample proportion

- As the sample size increases, the binomial distribution approaches a normal distribution.
- From the previous example:

$$\text{Standard deivation} = \sqrt{\frac{p(1-p)}{n}}$$

$$\text{Expected value} = E(\hat{P}) = p$$

$$sd = \sqrt{\frac{0.54 \times 0.46}{10}}$$

$$E(\hat{P}) = 0.54$$

$$sd = 0.17$$

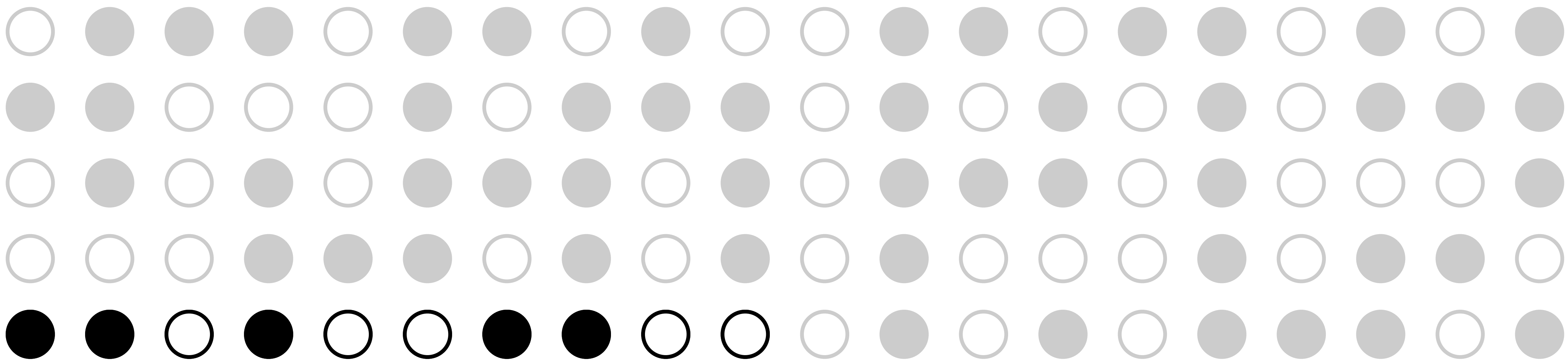
We can expect with around 68% certainty that the **sample proportion** will be within one standard deviation of the population proportion.

$$(0.37 < \hat{p} < 0.71)$$

We can expect with around 95% certainty that the **sample proportion** will be within two standard deviations of the sample proportion.

$$(0.20 < \hat{p} < 0.88)$$

Sample proportions



$$\hat{p} = \frac{\text{number of positive outcomes in sample}}{\text{sample size}}$$

$$\hat{p} = \frac{5}{10} = 0.5$$

What is the uncertainty of any estimates of the population proportion p?

What sample size is needed to be confident of correctly estimating p?

Confidence intervals

- Actually the **point estimate** of the sample proportion was 0.5.

$$\hat{p} = 0.5$$

$$sd = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$sd = \sqrt{\frac{0.5 \times 0.5}{10}}$$

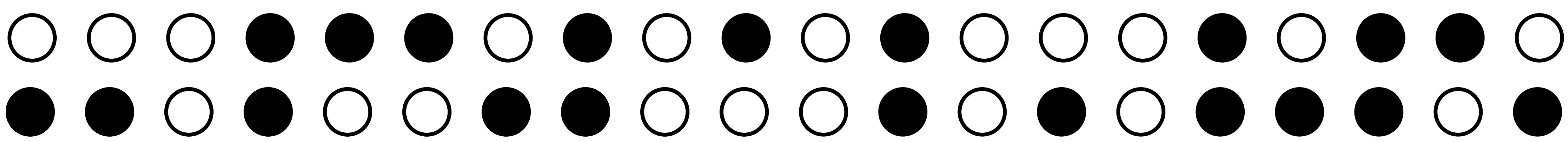
$$sd = 0.16$$

We can expect with about 68% certainty that the **population proportion** is within one standard deviation of the sample proportion. ($0.34 < p < 0.66$)

We can expect with about 95% certainty that the **population proportion** is within two standard deviations of the sample proportion. ($0.18 < p < 0.82$)

Margin of error

- The distance between the sample estimate and the end-points of the confidence interval is called the **margin of error**.
- To reduce the margin of error, the sample size needs to be increased.
- From a sample of 10, the margin of error at 95% confidence was ~0.32.
- To half the margin of error, the sample size should be four times greater.



Margin of error: $M \approx 2\sqrt{\dfrac{0.5 \times 0.5}{40}} \approx 0.16$

Margin of error

- The multiplier of the standard deviation needs to be found from the inverse normal distribution.
- For a 90% confidence:
- Find the value of z that has 95% of values below it.

80% : z=1.28

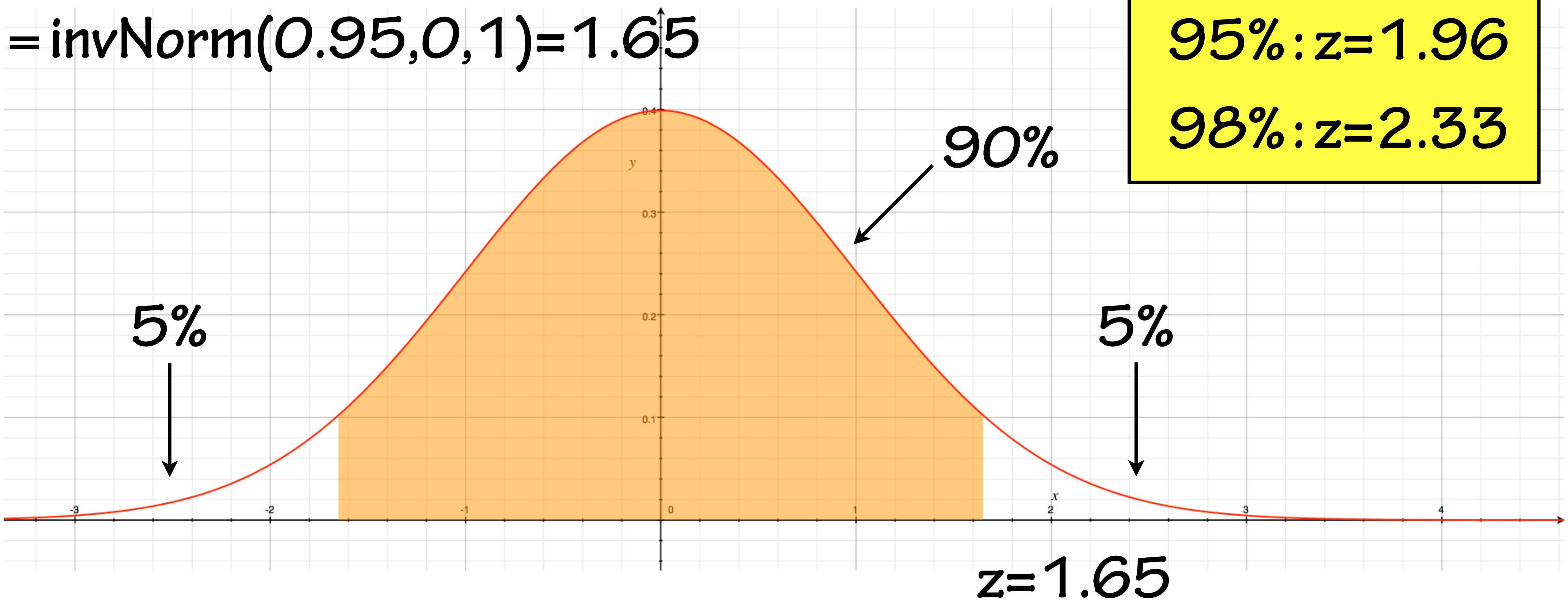
90% : z=1.65

95% : z=1.96

98% : z=2.33

90% confidence interval: Pr(Z>z)=95%

$z = \text{invNorm}(0.95,0,1)=1.65$



Margin of error

- A survey is to be taken of voters to find the proportion that have not yet decided on who they are voting for.
- How many people need to be surveyed for a 2% or 5% margin of error in the results with 95% confidence?
- Firstly, the sample proportion $\hat{p}$ must be estimated from prior data or a quick survey.

Assume that $\hat{p}$ is around 0.35 from preliminary data

$$0.02 = 1.96\sqrt{\frac{0.35\times 0.65}{n}} \qquad \left(\frac{0.02}{1.96}\right)^2 = \frac{0.35\times 0.65}{n}$$

$$n == \frac{0.35\times 0.65}{\left(\frac{0.02}{1.96}\right)^2}$$

$n = 2185$ (for 2% margin of error)

$n = 350$ (for 5% margin of error)

VCEMATHSMETHODS.COM

MATHS METHODS

PODCAST